User-Centered Design for Widely-Distributed Scientific Communities

Eckehard Doerry[1], Sarah Douglas[1], Arthur E. Kirkpatrick[1], Monte Westerfield[2]

[1]Computer Science Dept.   [2]Institute of Neuroscience
University of Oregon
Eugene, OR 97403
eck@cs.uoregon.edu

1.0 INTRODUCTION

The technology of the World Wide Web (WWW) provides a revolutionary means for dissemination of scientific information. For the first time, scientists have 24 hour, low-cost international access to central repositories of research data without the need for specialized client-side software. In particular, biological researchers have exploited the power of the Web to create a diverse range of bioscience resources, including several web-accessible relational databases (e.g., Mouse Genome Database [1], the Human Genome Database [2], the *C. Elegans* database [3], the Genome Sequence Database [4], and FlyBase [5]). There is also a growing number of other Web sites serving static HTML documents; by spring of 1996 there were 26 different Web sites for 15 different species, and the number of sites is increasing at a rate of about one every three to four months.

The biologists and computer scientists constructing these web sites presume that these web-accessible resources will aid scientific discovery through more timely, widespread access to better integrated research information. Although the WWW has made this information *physically* accessible to scientists, it is unclear whether it will be *cognitively* accessible. Busy scientists want useful, accurate, complete and up-to-date information without needing to learn and use a complex user interface. Will they be able to find answers to their questions without resorting to powerful but complex query languages like SQL? Will they be able to get their answers quickly without working through endless hierarchies of useless pages? Accessibility depends upon *usability* , and usability is critically related to productivity [6].

Designing a usable interface is challenging. First, the interface design must observe sound principles of graphic arts and psychology; it must have a functional layout, recognizable icons, and consistent interaction styles. Good design will reduce the time required to access desired information by minimizing misconceptions, mistakes and confusion in the search process. Second, it must incorporate a deep understanding of what information the scientist needs in the immediate context of his or her tasks and activities. In other words, it must present information using language and conceptual models understood by the scientist.

We have created a Web-based biological database for the zebrafish research community. The success of our project and the achievement of true accessibility and productivity depend upon designing, developing, and implementing with a user-centered approach. In taking this approach, we believe we are relatively unique among designers of web-accessible data resources. Here, we describe our use of this method and the unique challenges raised in applying it to a widely-distributed population of research scientists.

2.0 OVERVIEW OF THE ZEBRAFISH INFORMATION NETWORK (ZFIN) PROJECT

Researchers using zebrafish to study basic biology, like genetics and development, are distributed among more

than 100 laboratories in 28 countries. The zebrafish database project evolved out of our earlier Web site [7], which makes available (in static HTML documents) information on researchers and labs, a bibliography of publications relevant to the zebrafish research community, photos illustrating zebrafish developmental stages, and descriptions of laboratory methods, mutant lines, and the genetic map. The home page was accessed over 20,000 times in its first year.

Due to the exponential increase in information in this research area and the resulting need for more powerful methods of organizing and accessing these data, the zebrafish research community mandated extension of the original Web server to create a WWW accessible multimedia relational database known as the Zebrafish Information Network (ZFIN).

3.0 USER-CENTERED DESIGN

User-centered design focuses on the ultimate usefulness and usability of an interactive software product by assessing the requirements and specifications of the product from the user's point of view, the user's interactive behavior when using the software, and the context of its use. We have found that the only way to realize this vision is to relocate the entire design process to the user's domain. Rather than integrating biological expertise into a traditional software design effort, our aim has been to integrate computer support into the everyday situated work activities of research geneticists. Achieving this goal required the software design team to essentially become adjunct members of a genetics research lab in order to understand the overall scientific process, the role of information access in that process, and the relationships that exist between individual researchers, laboratory groups, and the research community as a whole.

We began by forging a participatory design team which included both biologists and computer scientists. We believe that this collaboration is key to the success of our project. Both domains are extremely specialized; creating an abstract data model for the domain, formally articulating the research processes that users are engaged in, and structuring of interface actions to match information seeking tasks all require both biological and technical expertise to achieve efficiency and usability.

Designing for a Distributed Scientific Community

Applying the user-centered methodology to a widely-distributed scientific community introduces a number of unique challenges. We identify three dimensions of difficulty: the heterogeneous nature of a scientific community, the lack of direct access to a significant portion of the community, and the technical challenges of interface design in the WWW environment. In this paper, we focus attention on the first two categories; technical challenges and our solutions to them are discussed elsewhere [8].

Scientific Communities

To date, most participatory design efforts have been undertaken within large companies where the target users are workers; domain experts are typically assigned to the design team by management. In contrast, a scientific community consists of loosely connected groups of relatively independent knowledge workers [9], each working on a particular aspect of the same general problem. Table 1 exposes the special challenges encountered in designing for a scientific community by comparison to a more traditional corporate context:

	Scientific Contexts	Corporate Contexts
Domain Knowledge	Domain knowledge is extensive, diverse and very difficult to acquire (e.g. developmental genetics). Domain knowledge is constantly changing as the field evolves.	Relevant domain entities and procedures are generally well-established and stable.
Information Flow	Many forms of formal and informal information exchange exist. Formal information exchange (e.g. publication) is highly institutionalized to guarantee accuracy and proper attribution; informal information exchanges abound and are ever-changing as workers migrate between labs, people enter and leave the community, and projects change.	Well-defined channels of information exchange have been defined by management. Informal means of information exchange exist as well, but are generally stable once established.
Organization	Labs form the basic social group, with each lab headed by a principle scientist and containing other research scientists, post-docs and doctoral students. Labs are loosely-connected into a global scientific community.	Work groups form the basic social group, forming the leaves of a hierarchical corporate structure. A uniform set of working procedures for each group is dictated from above; work groups are ultimately directed by a single executive power.
Culture	Labs and individual scientists within them are both cooperative and competitive. Progress of the discipline depends on sharing research results; success of the individual depends on attribution of results to the individual.	Success of the individual is intimately tied to the success of the work group and, more generally, to the success of the corporation. Unfettered information flow within the company is encouraged.

Table 1: Differences between scientific and corporate design contexts.

In general, the differences highlighted in Table 1 emphasize the independent, heterogeneous nature of scientific communities. This has forced us to modify the user-centered design process in a number of ways. Rather than studying and designing for a single, representative group of users, we have had to come to grips with the fact that no such group exists in a diverse scientific community. Although, from a practical perspective, intensive ethnographic user analysis must be reserved for one or two highly-accessible groups of users, we have invested much effort in sharing the resulting insights with the remainder of the community, generalizing our formal models of domain activities and data to encompass the differences that exist within the community. A simple example is the representation of developmental time within our abstract data model. Our initial analysis characterized the developmental age of an embryo in terms of a closed set of named developmental "stages" defined by the zebrafish community. However, later analysis revealed that different laboratories may have different conventions for specifying developmental age, e.g., by giving the age of the embryo in hours or, more coarsely, in days. The data model had to be generalized to accommodate these different metrics while still supporting efficient indexing and retrieval of data.

A related difficulty is the changing structure of domain knowledge in scientific domains. In most corporate contexts, the domain data model is relatively stable. Though information content may come and go, the *kinds*

motivating researchers to keep information to themselves. A good solution to this dilemma must maintain proper accreditation of work while making new results available soon after they are discovered.

Widely-distributed User Community

The widely-distributed nature of scientific communities interferes directly with one of the basic tenets of user-centered design, namely the immersive involvement of designers in the everyday workings of the community. Accordingly, a primary challenge was to find ways of adapting the user-centered design process to a distributed design context, balancing the need for intensive ethnographic analysis with the need to expose and accommodate the diverse requirements of the entire community of users. As discussed in the following section, we addressed this difficulty using a two-pronged approach, capitalizing on professional meetings attended by most community members whenever possible, while utilizing the WWW to disseminate information and collect feedback on the nascent data model and the information access tasks to be supported by the proposed scientific database.

User-Centered Design: Process

Our design process (Figure 1) follows the basic steps of user-centered design [10].
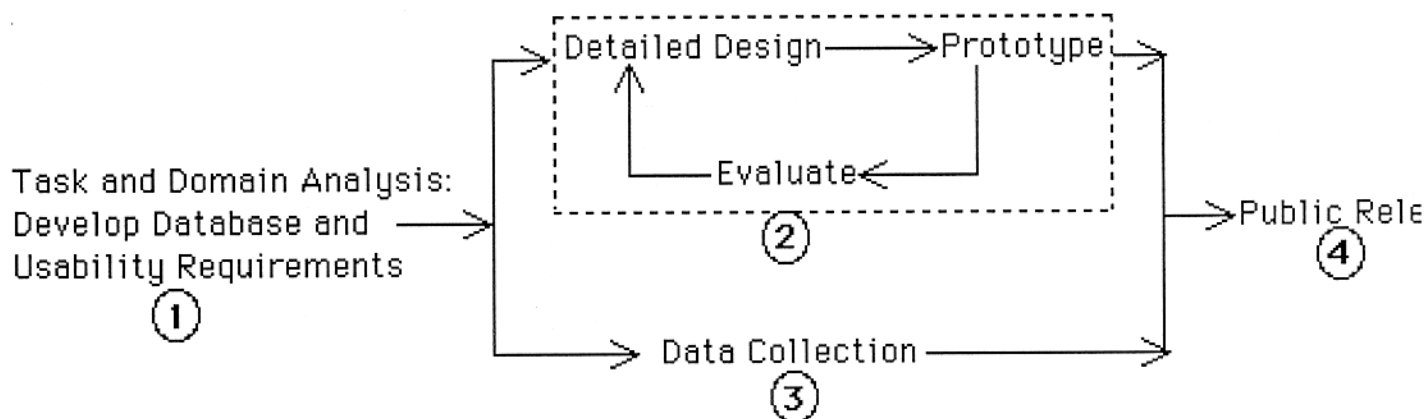

Figure 1: Steps in the user-centered design process.

Step 1: Develop database and usability requirements

The initial step in our design process is to conduct domain and task analyses. The goal of this step is to produce both database information and interactive system specifications. However, the biological domain is extremely specialized, making this analysis very difficult. Development of the abstract data model, the nomenclature used to label user interface components, and the structuring of interface actions into information seeking tasks all require deep knowledge of the domain to achieve efficiency and usability. Accordingly, we began the design process by forging a participatory design team [10, 11] which includes both biologists and computer scientists. We believe that this collaboration is key to the success of our project, not only because it provides domain and task knowledge, but also because direct involvement of biologists gives them a stake in the success of the project.

During Step 1 (Figure 1) we used primarily ethnographic methods [12], including interviews with zebrafish scientists, reading journal articles, attending research talks and lab meetings, and participant-observer activities such as helping to customize the specialized software used by one of the labs. In other words we tried to "go native" in the zebrafish community. We also used questionnaires to gather design input from scientists around the world, distributing them at workshops, and via the original Zebrafish Web site, which contains documents on the development of the database project along with a brief list of the types of information we expect to include. We solicited feedback from our users about their satisfaction with the current HTML-based Web site and integrated their requests for enhanced functionality and information into the design of the new database.

The goal of this immersion in the working world of zebrafish scientists was to understand the context of their everyday work activities and its relationship to the proposed WWW database.

In addition to these ethnographic methods, we looked extensively at other web-accessible biological database sites to evaluate their information content and user interfaces; we videotaped our own zebrafish scientists doing simple information retrieval tasks to pinpoint their confusions with the user interfaces and information models at these other sites. In this respect, we have found the WWW to be an easily accessible means of drawing on the design experience of others, a critical resource in an area where design improvements are typically incremental and based on real-world experience.

These domain and task analyses produced specifications for the database information and the user interface. Database information specifications were captured in a data model document intelligible to both computer scientists and biologists, containing descriptions of database entities, attributes, and relationships, as well as examples of situations of use for various pieces of information. The data model serves as the blueprint for database implementation and offers zebrafish biologists a concise overview of database contents. The current design for the database is complex and incorporates 21 major classes of information, most of which are highly interconnected.

Step 1 (Figure 1) also produced specifications for the interactive system. In particular, functional and performance requirements of the system were determined *as seen from the user's point of view* [13]. The first of these, functional requirements, describes what the system should do. Our functional requirements include:

- Must provide a security mechanism to ensure that only authorized users submit data.
- Must guarantee reliability and completeness of the data, especially because much of it is user supplied.
- Must have a mechanism to distinguish published data from unpublished or pre-published data.
- Must allow submission of commonly published image types, including annotations.
- Must provide color reliability and reasonable resolution for images so that information is accurate enough to make science-based decisions.
- Must be accessible from multiple platforms, including older machines, to support universal access to the database.
- May need to access other databases concurrently with ZFIN.

Performance requirements, the second type of requirements, state how well the system should perform from the user's point of view. Thus, performance requirements define criteria for evaluating the actual usability of the resulting design used in Step 2 (Figure 1) of the design process. Our performance requirements include:

- Must be easy to learn. We expect user interactions with the database (retrievals or submissions) to be relatively infrequent, and thus, a typical scientist might forget how to use the interface between uses. The interface must be learned as you go (no separate manual) and must provide extensive on-line help.
- Must be fast enough to satisfy most commonly asked questions within a 10 minute session.
- Must provide enough feedback that most searches will find results with three rounds of querying.
- Must keep the user apprised of progress during the data submission process, allowing the user to "undo" a submission during any step.

It was apparent from our requirements study that we needed to offer a more usable interface than SQL. Although SQL is extremely expressive, allowing extraction of very complex database relations, it is practically impossible for non-database professionals to learn and use [14]. Thus, a primary challenge in the design of the ZFIN user interface was to determine *in advance* a subset of queries which would satisfy the needs of most users and to create a very simple interface for expressing queries in this subset. This required an extensive understanding of the domain and tasks.

Step 2: Iterate detailed design process

Step 2 (Figure 1) is the heart of usability engineering methods [15]. After developing the information and usability requirements, we moved into the iterative refinement phase of the user-centered design process to

design and implement the user interface. In contrast to the waterfall approach used by traditional software engineers, this technique relies on rapid cycles of design, prototype implementation, and evaluation with real users to generate the final product.

Rather than implementing the entire database at once, we initially selected a subset of the database information (i.e. data types) to take through Steps 2-4 of the design process. This was done primarily for pragmatic reasons. Some information is of higher priority to the research community or more mature and complete; staggering development of various data classes allowed us to make useful information rapidly available. In addition, focusing attention on just a few types of information at a time proved to be an effective means of managing the complexity of the design.

Each prototype was immediately evaluated by selecting pairs of zebrafish scientists and giving them typical data retrieval and submission tasks to perform. Videotaping these sessions allowed us to analyze the amount of time required, misconceptions encountered, and other problems with the interface. We evaluated their performance against the usability requirements developed in Step 1 (Figure 1). Details of how to conduct this type of performance analysis can be found in [16]. We used insights gained from this analysis to shape subsequent prototypes in the iterative design cycle.

When we were satisfied with the usability of a prototype, it was made available to 10 zebrafish scientists (acting as beta testers) through the WWW; access to the prototypes was limited to these testers. Feedback on the prototype was gathered using a series of short, on-line questionnaires and by including a "comment" button on each screen of the prototype, allowing users to easily generate an email message to the developers. In addition, the system recorded the screens traversed by each user, allowing us to identify areas of particular interest and to expose problems with the interface. At the end of the beta testing period, we also interviewed our testers to discover any other problems. The prototypes were also demonstrated at special sessions during professional meetings, with feedback gathered orally and by distributing short questionnaires.

As a result of this participatory process, we have developed a committed group of "co-developers" in the user community that is highly representative of the global community as a whole.

Steps 3 and 4: Data Collection and Public Release

Because data collection (Step 3) is not an intrinsic part of the user-centered design process, it will not be addressed here, in favor of a more focused discussion of usability and interface design issues. Step 4, public release of the database, is an important part of the user-centered design process rather than its abrupt termination; usability analysis will continue indefinitely, allowing the system to evolve to meet the changing demands of users. The commentary forms for gathering user feedback mentioned earlier in the context of beta testing remain available in the public release. We are also recording (anonymously) the sequence of screens visited by each user and the total number of visits to each screen in an effort to determine common usage patterns and to expose areas of confusion. Finally, we are planning to conduct extensive periodic user surveys, interviewing a sample of randomly selected registered ZFIN users to assess usage patterns, good and bad features of the Web site, and interest in future information support.

5.0 CONCLUSIONS

Rapidly expanding access to the WWW holds incredible promise for increased data sharing and collaboration within widely-distributed research communities. Web-accessible databases will make available a much broader range of data than printed media, including multimedia data types and information that, although useful, might never be formally published; new findings can be made available almost instantly, rather than being delayed for months by a lengthy editorial and printing process.

There are a number of challenging obstacles to such universal accessibility. First, scientific domains are unusually complex, with domain ontologies that evolve with the expanding frontiers of the discipline; the kinds of information accepted as "data" and the research techniques that generate this information change over time. Consequently, interface design for a research database is much more demanding than for stable, single-use

databases.

Our experience with this project demonstrates that user-centered design can be used to manage domain complexity and generate meaningful usability requirements, by focusing attention on the real-world work activities (i.e. research processes) of users and on the ways in which access to various kinds of information (data) contributes to these activities. It also demonstrates that, by using the WWW as a medium for both the design process and the design itself, user-centered techniques can be adapted to contexts in which the user population is widely-distributed and loosely-organized.

We have focused our work to date on maximizing the accessibility of data contained in the ZFIN database, applying user-centered techniques to streamline individual data manipulations. In the future, we plan to expand our focus to support the full research process in which database access is embedded by developing tools to support database-centered interaction among widely-distributed members of the research community. Examples include mechanisms for collaborative shared access to the database, interactive discussion forums, and viewer commentary appended to specific data records. We envision ZFIN as the cornerstone of a virtual community of research scientists linked via the WWW, working together, and sharing a common set of data.

Acknowledgments

References

1. Jackson Laboratory (1996), Mouse Genome Database, (World Wide Web URL http://www.informatics.jax.org/): Bar Harbor, Maine.

2. Johns Hopkins School of Medicine (1996), Genome Database (GDB), (World Wide Web URL http://gdbwww.gdb.org/): Baltimore, MD.

3. Genome Informatics Group (1996), ACEDB, US Department of Agricultur (World Wide Web URL http://probe.nalusda.gov:8300/cgi-bin/query?dbname=acedb).: Beltsville, MD.

4. National Center for Genome Resources (1996), Genome Sequence Database (GSDB), (World Wide Wed URL http://www.ncgr.org/gsdb/): Santa Fe, NM.

5. Harvard Medical School (1996), FlyBase, (World Wide Web URL http://cbbridges.harvard.edu:7081/), Harvard Medical School: Cambridge, MA.

6. Landauer, T.K. (1995), The trouble with computers. Cambridge, MA: MIT Press.

7. University of Oregon Institute of Neuroscience (1996), The FISH Net, (World Wide Web URL http://zfish.uoregon.edu): Eugene, OR.

8. Doerry, E., et al. (1997), Moving beyond HTML to Create a Multimedia Database with User-Centered Design: A Case Study of a Biological Database, University of Oregon Technical Report CIS-TR-97-02: Eugene, OR.

9. Crane, D. (1972), Invisible Colleges. Chicago: University of Chicago Press.

10. Schuler, D. and A. Namioka (1993), Participatory design: Principles and practices. Hillsdale, NJ: Lawrence Erlbaum Associates.

11. Greenbaum, J. and M. Kyng (1991), Design at work: Cooperative design of computer systems. Hillsdale, NJ: Lawrence Erlbaum Associates.

12. Blomberg, J., et al. (1993), Ethnographic field methods and their relation to design, in Participatory Design: Principles and Practices, D. Schuler and A. Namioka, Editors. Lawrence Erlbaum Associates: Hillsdale, NJ. p. 123-155.

13. Newman, W.M. and M.G. Lamming (1995), Interactive systems design. Wokingham, England: Addison-Wesley.

14. Greene, S.L., S.L. Gomez, and S.J. Devlin (1986), A cognitive analysis of database query production. In Proceedings of the Human Factors Society. Santa Monica, CA: Human Factors Society.

15. Nielsen, J. (1993), Usability Engineering. Boston: Academic Press.

16. Douglas, S.A. (1995), Conversation analysis and human-computer interaction design, in Social and Interactional Dimensions of Human-Computer Interfaces, P.J. Thomas, Editor. Cambridge University Press. p. 184-203.